DOCUMENT RESUME

ED 088 924 TM 003 481

AUTHOR Thompson, Charles W. W.; Bath, Gustave J.
TITLE The Administrative Experiment: A Special Case of Field Testing or Evaluation.
PUB DATE [71]
NOTE 34p.

EDRS PRICE MF-$0.75 HC-$1.85
DESCRIPTORS *Administrative Personnel; Administrator Role; *Evaluation Methods; *Field Studies; Research; Research Design; *Researchers; *Research Methodology
IDENTIFIERS *Administrative Experimentation

ABSTRACT

This paper describes administrative experimentation--a particular strategy and method of field testing or evaluation. Administrative experimentation can best be described as a specialized form of field testing, field experimentation or evaluation which emphasizes the dual role of administrator and experimenter. This method is first defined in terms of its relation to other methods, and its potential strengths and limitations. Second, the method itself is described in some detail. Finally, case studies of its use are presented to provide a basis for evaluating the usefulness of the method. (Author/PTP)

ED 088924

TM 003 481

THE ADMINISTRATIVE EXPERIMENT: A SPECIAL CASE OF FIELD TESTING OR EVALUATION +

Charles W.N. Thompson and Gustave J. Roth ++

Northwestern University

Evanston, Illinois

ABSTRACT

This paper describes a particular strategy and method of field testing or evaluation — administrative experimentation. The method is first defined in terms of its relation to other methods, and its potential strengths and limitations. Second, the method itself is described in some detail. Finally, case studies of its use are presented to provide a basis for evaluating the usefulness of the method.

+ This paper was supported in part by PHS Grant CH00454-02.

++ Portions of this paper were presented at the APA 1971 Session on September 6, 1971.

72-1-20

ED 088924
TM 003 431

RELATION AMONG FIELD TESTING, EVALUATION AND EXPERIMENTATION

Definitions

A wide variety of activities are undertaken to determine the effects of introducing changes under field conditions. The persons engaged in these activities may range from naive and casual observers to sophisticated and meticulous basic researchers, and their purposes may range from a passive acceptance of stimuli to an understanding of a particular phenomenon in terms of a rigorously defined test of a well developed theory. The terms field testing, evaluation and experiment are used, with various modifiers and meanings, by different people to describe these kinds of activities, and they generally serve to set these activities apart from those which are not concerned with (introduced) changes, e.g., some kinds of base line studies or one-time observations, or are not concerned with field conditions, e.g., abstract speculation or laboratory experiments, or are not concerned with "determining" the effects, e.g., the incurious, non-scientific or uninvolved.

Not only in the course of describing the results of particular activities but also more directly much has been written to explain what these terms mean. Evaluation (and evaluative research, program demonstration, action research, organizational development, etc.) is, perhaps, the broadest of these terms, and may be defined as a "concern with both information on the outcome of programs and judgments regarding the desirability or value of programs." (Caro, 1971, p.2) This goal oriented sense is emphasized by Suchman (1967, p. 32) in distinction from "evaluative research" as one means for reaching that goal. Testing and field testing are similarly broad terms often defined, such as in contracts, program documents, and specifications, by descriptions of activities involving operation, measurement

and comparison, or, in some cases, with considerable effort for definition (Kerke, 1963). Experiment (natural, field, quasi, administrative, etc.) rounds out this set of broad terms and, of the three, probably provokes the most learned arguments, but it may be usefully defined as the "basic approach (of the scientist)...to intrude and interrupt, to make a change and see what happens." (Campbell, 1967, p. 258)

The differential preferences in the use of these terms by people from different disciplines or vocations may reflect their purposes or objectives either in the particular case or as a matter of their general orientation. See, for example, Morgan (1971) or Erdmann and Neal (1971). Because purpose is so person-specific, it may be desirable to concentrate on the methods used to examine the several meanings of these terms and their relation to other similar or related terms preliminary to a more detailed examination of a particular method — administrative experimentation. For this purpose, three dimensions will be proposed.

Exploratory ←→ A Priori

The first dimension will be called EXPLORATORY ←→ A PRIORI, and the dimension may be described as variations in "the degree to which the researcher (tester, evaluator) predetermines the effect of new data (obtained from the phenomena) on his results." This dimension characteristically distinguishes the early exploratory stages of the research (or engineering or any other) process from the later stages of replication and confirmation. In the early stages uncertainty may exist not only with respect to the relationship among the variables of interest but also with respect to a wide variety of parametric conditions and the methods most appropriate for examining them. Often it is not economical, even if it is feasible, to attempt to specify in great detail what information will be obtained and how it will be interpreted; instead, as the new data is gathered, it is examined in various ways

(a posteriori manipulation) and a variety of preliminary hypotheses may be generated. Later, as theory and data build up, the general uncertainty may diminish, and it may be appropriate to examine some particular hypothesis with great care, controlling for alternate hypotheses (parameters), arriving at a position where the researcher has predetermined (defined beforehand) exactly how his new data will affect his results.

For example, the researcher looking for, or expecting to find, new phenomena, e.g., an anthropologist's first visit to an obscure primitive society, may start with a lot of information on the various relations that may exist and on the methods for examining them, but what he observes, at least initially, will only provide a plausible or illustrative level of understanding of the phenomena. His purpose will be to develop preliminary hypotheses, and this may be a valuable scientific accomplishment. At the other extreme, the researcher who wishes to confirm a well-developed theory which has already been subjected to evaluation, e.g., an experimental psychologist replicating a particular experiment, may follow a detailed research design which specifies not only all of the conditions of the observation but also what the results will be. His purpose will be to assure that the description or explanation which he has is supported by compelling evidence.

Similarly, in the early stages of design, the testing of a breadboard may be to establish whether a particular circuit will work at all, or how variations in its arrangements will affect the results. Much later, in acceptance testing, a detailed test design will be used, and this will include a statement of the results expected. With respect to this dimension, evaluation and field testing, and other categories of research, may appear anywhere. Caro (1971, pp. 2-3) discusses approaches to evaluation which appear to span this dimension, with "evaluative research" tending to be

closer to hypothesis testing. Field testing would appear somewhere in the middle, but it is proposed that an examination of reports of specific tests would show a considerable spread. For a related discussion of system design, see Knowles, Burger, Mitchell, Hanifan and Wullock (1969).

Arguments or judgments as to the comparative value or "correctness" of one type of activity or the other may contribute little without an understanding of the purpose of the researcher, and a recognition of the relevant state-of-the-art which he is faced with initially. It is only when the researcher chooses a strategy inappropriate to his purpose (and/or the state-of-the-art), or misdescribes what he did, that a criticism becomes significant. To illustrate how a variety of activities relate to this dimension, reference may be made to Figure 1. The relative position on the table is representative, and a particular activity may vary considerably. With a more precise statement

.....

Figure 1 about here

.....

of the dimension, and a specification of various parametric conditions (researcher's purpose, period of activity, effect of others, etc.), it should be possible to provide a more descriptive statement which could then be "tested" by relating the use of the term to the way in which the activity was carried out.

Normative ⟷ Empirical

The second dimension will be called NORMATIVE ⟷ EMPIRICAL, and the dimension may be described as variations in "the degree to which the researcher (tester, evaluator) obtains new data directly from the phenomena." This dimension, illustrated in Figure 2, is similar to those which are represented

.....

Figure 2 about here

.....

by distinctions such as, between levels of abstraction, between theorizing and study or experiment, and between "ought to be" and "is." The critical characteristic which the dimension concerns itself with is the degree of mediation (modification, transformation, etc.) between the real world phenomena and the researcher's head. This mediation may be introduced during direct observation by the process of perception itself, and further mediated by interaction with other information (or data) with the passage of time, or form of recording, or merely by the complexity and amount of information obtained.

Direct observation of simple, well understood phenomena, such as the [illegible] would illustrate one extreme; the manipulation of a mathematical expression with no explicit reference to any real world phenomena might serve to mark the other extreme.

On this dimension, laboratory and field experiments (and studies) are not distinguished; both involve direct interaction with the real world phenomena, in the sense that a laboratory subject (or equipment) exists in the real world. It is as the "complexity" of the observation increases, as measured by the purpose (and/or viewpoint) of the observer, that the mediation may increase; and, to the extent field observations are more complex, they may be distinguished from laboratory experiments. Similarly, field testing is usually thought of as involving a high degree of direct observation of the real world phenomena, although complexity may introduce considerable mediation. Where evaluation is based on direct observation it would be similarly located on this dimension; where evaluation is based on reports or

other mediating sources, and modified by information, such as value judgments, prior opinions, historical data, etc., it would be less direct.

In this general sense of mediation, the reports of others may be located on this dimension. Thus, reports of observations and experiments would, in form and content, be closer (less mediated) to a direct observation than a theory paper; this mediation is analogous to that introduced by the reports of observers in a large evaluation or testing program.

Study ⟷ Experiment

The third dimension, and the last one proposed here, will be called STUDY ⟷ EXPERIMENT, and the dimension may be described as variations in "the degree to which the researcher manipulates the phenomena under observation." This dimension, illustrated in Figure 3, appears to be self-explanatory but subtends several critical problems in definition and execution. In the

.....

Figure 3 about here

.....

simplest case, a laboratory experiment, the researcher makes a change in one variable and observes the effect on another, e.g., he moves a magnet close to a paper upon which are scattered iron filings and observes the pattern they form. Here it is clear that the researcher manipulated (intervened in, changed, etc.) some part of the real world, i.e., experimented.

If, instead, the change, the moving of the magnet, were done by an assistant, under the researcher's direction, the sense of experiment would be compromised only to the extent the researcher's direction was not carried out, and this may, in some cases, be significant. Further extensions would include cases where the direction of the researcher becomes, progressively,

less of an influence on the actual manipulation, e.g., the manipulator is an administrator who "agrees" to make the change, or has informed the researcher that he is making a change which the researcher would have directed if he could have done so. A second variation, the "natural experiment," appears where the change which the researcher desires to bring about (manipulate) may occur without his direct intervention, and he can accomplish the same purpose in his choice of what to observe, or when, or how. This is a particularly appropriate choice where he cannot, in fact, make the change desired, e.g., in making an astronomical observation. A third variation which presents a different problem in manipulation, arises where the researcher does not want to influence, such as to avoid biasing or contaminating the events he is observing; this particularly arises in the study of an ongoing organization.

All three of these variations may occur in field experiments, field testing, and evaluation, particularly if a large and complex change is to be introduced. Seldom does the researcher directly introduce the changes desired; there may be changes which he cannot bring about; and there may be changes which he does not want but cannot prevent.

Comments

One purpose which these dimensions serve is to focus attention on the similarities and differences among a wide variety of activities. These may reflect the purposes of the researcher (evaluator, tester), the initial state of his knowledge (how early or late he is in the process of reducing his uncertainty about the phenomena), to what extent the new data he obtains from the phenomena is mediated, by choice or circumstance, and to what extent he manipulates the phenomena, whether he does or does not intend to do so.

A second purpose is to demonstrate that the wide variety of activities that may be employed to improve one's understanding of "what is going on" have some underlying characteristics which may provide a more useful basis for classification than ethnocentric categories of "good" and "bad" methods. It would seem clear, for example, on the EXPLORATORY ⟷ A PRIORI Dimension, that both exploratory and explanatory research may be useful, i.e., "good," under appropriate circumstances. An examination of "good" and "bad" research on this dimension would, it is suggested, disclose that "good" research is that in which (at least) the researcher bases his choice of method on the degree of his initial uncertainty, and is careful to disclose the accompanying degree of uncertainty in his results. See, for example, Liebow (1967, pp. 10-12).

In a similar manner, it can be proposed that the other two dimensions illustrate the importance of choosing the method most appropriate to the purpose of the researcher, and the circumstances which are presented. However, where the purpose is to improve one's understanding of "what is going in the real world," both dimensions suggest an order of preference and the reasons underlying that preference. In the NORMATIVE ⟷ EMPIRICAL Dimension, the preference would be towards minimizing the mediation (distortion, "noise," error, etc.) by preferring methods toward the right hand side. If circumstances "force" the researcher to use in part, or all, methods with more opportunity for the introduction of mediation, he should then direct his attention to the potential effect of this mediation, and he should make sure that his results properly report this effort. In the STUDY ⟷ EXPERIMENT Dimension, there are conflicting preferences in the choice of method. Those on the right hand side make it possible for the researcher to control the time and form of the change he is interested in; as he chooses methods further away, he is less able to ensure his control, or even the presence, of the variable he is interested in. The conflicting preference, to minimize

the manipulation of other variables, is most likely to be realized at the extreme left and right hand sides, with the least preference in the center of the chart. Where the choice of a "natural experiment" is not indicated, the researcher should direct his attention to maximizing his ability to control the variables he is interested in and avoiding or minimizing his manipulation (contamination) of other variables; and he should make sure his results properly report this effort.

ADMINISTRATIVE EXPERIMENTS

While the discussion above was often in terms of the "researcher," it should be clear that the comments were also directed to the "evaluator" or "tester" or "experimenter." In all of these cases, the subject under investigation could be a purely physical phenomenon or system, but the comments could be applied to a man or a man-machine system, and, because, the concern of the present paper is with the latter, the discussion from here on will be so directed.

Administrative experiments may be classified as a type of field experiment, and, as such, may be quite similar to field testing and evaluation. Reference to the tables will illustrate this. As a "field experiment," the method suggests an intentional manipulation of the phenomena in the real world, with care to avoid introducing unwanted manipulations (STUDY ⟷ EXPERIMENT Dimension), and a strong preference for obtaining the data as directly as possible from the phenomena (NORMATIVE ⟷ EMPIRICAL Dimension). The method (as described here) also proposes to predetermine the effect of new data as much as the circumstances (the degree of uncertainty) at the initial point of the research allows, but to adjust to those circumstances if required and to carefully report the effect on the results of that adjustment (EXPLORATORY ⟷ A PRIORI Dimension).

The use of the term "administrative" reflects the fact that many, if not most, of the interesting changes in the real world require that the manipulation be done by the person who controls the variables of interest. While researchers have been able to obtain the cooperation of administrators, it was argued by Campbell (1967, pp. 258-259) that this was not enough, as follows:

> A social science that is to deal with important problems must tie in with people who are doing experimentation at more important levels, who have arbitrary power over more sustained interactions, larger groups of persons, and a wider range of settings. This power is held by the administrators, the people who are in charge of factories, training programs, and school systems; the people whose decisions (whether democratically or arbitrarily arrived at) are put into effect; the people who are the gatekeepers (democratically or dictatorially) for abrupt rather than gradual changes. Thus, if we are to have an experimental social science, the social scientist must develop a liaison with the people who have the power. It is not we but the administrators who have the experimental laboratories, through being at the site, if not the decision-making seat, when abrupt administrative policy changes are made. What we social scientists must do is to convince administrators of the necessity of keeping books on the experiments they make and organizing their record systems and publication practices so that they let us know what they have tried and how it came out.

This, of course, argues the contribution of the method to social science, but there are advantages (and disadvantages) for the administrator and those, in his organization, who decide to use the method.

The advantages to an administrator in applying the more formal methods of administrative experimentation to the changes which he has introduced have been outlined by Thompson (1969b, p.3) as follows:

1. He may use the methods as a framework for identifying the claims which appear in the literature, or elsewhere, and to

evaluate the bases of these claims and their application to his purposes.

2. He may use the methods in evaluating (or testing) the changes he institutes in his own management solutions to determine if the changes were what made the difference in results.

3. He may use the methods, and the record he keeps, to support his claim, such as to his own management, that his management solutions were, in fact, the basis of the results achieved.

It is perhaps self-evident for many, if not most, management problems that difficulty in coming up with a solution is not as frequently encountered as being presented with too many suggested alternative solutions. And,

> (f)or all the talk about scientific management and management principles, (the administrator) may find it difficult to understand why a solution works sometimes and does not work at others, and why there seem to be so many alternatives which appear to be clearly inconsistent (Thompson, 1967, p. 2).

This difficulty applies, also, to the administrator's claims with respect to his own innovations. There are certainly few things more frustrating than being right and not being believed. Through the methods of administrative experimentation, the administrator may anticipate, and meet, some, if not all, of the post hoc doubts (plausible rival hypotheses) which others may raise as limiting the credibility of his claim.

The administrator who is considering the use of the methods of administrative experimentation will be faced with a number of disadvantages. Campbell (1967, pp. 287-290) has called these "sources of resistance to experimentation and evaluation." These include the following: a) the additional cost of record keeping (and planning, and evaluating); b) the political vulnerability from having "hard facts" available concerning the effectiveness of the

administrator; c) the disappointment if the result of the experiment is that the change did not make a difference; d) the appearance of indecisiveness in decision making which a willingness to admit that a test is necessary (or desirable) suggests.

There are other advantages and disadvantages growing out of the differences between an administrator (or inhouse researcher) and a university (outside) researcher, and these have been discussed in a number of places. The administrator has the advantage of knowing a lot about the other factors (parameters) which may affect his experiment; he may be better able to manipulate variables (and he is the ultimate judge of the propriety of doing so); and he may have better access to records and other sources of information. On the other hand, he may be more tempted to sacrifice the experiment to ensure a "good" outcome; he may be less familiar with the results of the experiments or studies conducted by others, or of the variety of experimental methods available; and he may be less able to ensure the confidentiality necessary to obtain certain kinds of data.

METHOD

Preliminary Comments

There are some preliminary assumptions or "givens" in the method to be described which should be understood at the outset. While the basic methodological argument is more broadly applicable, the details can best be understood in view of these assumptions.

The first assumption is that the researcher (tester, evaluator) has some degree of preference for experimentation (as distinguished from a STUDY approach); this may be based on no more than opportunity, but there are certain advantages, as outlined by Chapanis (1959, pp. 148-149).

> When the scientist sets up an experiment he plans, controls and describes all of the circumstances surrounding his tests. Not only does this give him greater control over the course of events but it also enables him to set up conditions so that he can repeat the experiment if he wants to.
>
> . . . . . .
>
> One of the most important reasons for creating such artificial situations is that the experimenter can make an event happen at a certain time and place. This means, among other things, that he is prepared to make accurate observations of the event because he knows when to expect it. Another reason why experiments are done in an artificial setting is to allow the experimenter to control and manipulate the variables which might affect the outcome of his observations... Another important difference between an experiment and mere observation is that in an experiment the research man can systematically vary conditions and note the concomitant

> variations in results...This also means that the investigator can, if he wishes, try combinations which do not occur, or have not yet occurred in real life.

The second assumption is that the researcher (tester, evaluator) has some degree of preference for real world experiments (as distinguished from either a laboratory or NORMATIVE approach); this, again, may be based on no more than opportunity, but there are certain advantages, as outlined by Chapanis (1959, pp. 199-200).

> The most important argument for maximum realism is that we have the greatest confidence in the results of such experiments--that is, we have confidence in them if they are, or can be, done well. The greater the realism of our experiments, the more certain we are that they will tell us exactly what will happen in real-life situations.
>
> . . . . . .
>
> ...we cannot help having some residual doubts in applying laboratory data to real-life situations: Did we really think of everything important in the laboratory experiment? Did we consider all the relevant variables? Ultimately, our only validation comes from field trials, or real-life tests...Aside from the difficulties of controlling variables, there is another, quite different reason for making human engineering experiments as realistic as possible. The typical laboratory experiment is a highly artificial situation, which the subjects perceive as such. As a result, subjects do not behave the same in the laboratory as they would in a real-life situation.
>
> . . . . . .

> Another aspect of the artificiality of laboratory experiments is the fact that when you take a problem into the laboratory, the variables may not be the same as those in real life...At best, a laboratory experiment is only an approximation to life.

The third assumption is that the researcher (evaluator, tester) has some degree of freedom of choice of the problem. In the strictest sense, an "administrative experiment" is one in which the same person is both the administrator and the experimenter. As administrator, he chooses the "experiment" for reasons related directly, or indirectly, to his responsibilities as an administrator, and he is in a position to introduce the experimental condition because the variables are under his control; as an experimenter, he plans and carries out the experimental design to maximize the credibility of his results. But, even under these circumstances, he may not be a completely "free agent"; he will have regard for the objectives of the organization or even specific "experimental objectives" of others; he may depend upon the advice and participation of others in the design and carrying out of the experiment itself. Where others have a significant input, such as in the choice of the problem, the administrator-experimenter may still be able to use the method outlined here if he has sufficient latitude to reconcile his two functions, and some choice in how to carry them out.

The fourth, and last, assumption is that the researcher (tester, evaluator) will agree that the "steps" in the method are iterative and interactive but without ordinality. The sequence of presentation is purely for convenience; one can begin with any or all of the steps, progressing iteratively by drawing upon what one has done with any of the other steps, and ending with an integration of what one has done with all of the steps.

Choosing the problem

In principle, any variable(s) which the researcher (administrator) can manipulate may provide the basis for an experiment, and there are no rigid rules for selection. There are, however, some considerations which may be useful, the first of which is that the researcher should check the problem he has initially chosen against all of the "steps" in the process before he makes a commitment to the problem.

An obvious starting point would be a "change" which the administrator is going to make anyway, or is considering making. It may be based on experience, or some current problem, or from the literature, or suggested by others. It can, and perhaps should be, a change which the administrator thinks will "work." It should not be trivial in the sense that "everyone" already knows the answer, or in the sense that the effect is so minor or unimportant that "no one" will care to know the results. On the other hand, it should not be so large and complex (cosmic) that a credible solution cannot be achieved within a reasonable time and with reasonable resources.

In the language of experiments, the variable (or variables) which is to be manipulated is called the independent variable. To be of some practical value, it should not only be a variable which can be manipulated for purposes of the experiment but also a variable which would be manipulated by an administrator seeking the effect which the experiment has established. Similarly, the effect predicted is in terms of changes in the value of the dependent variable (or variables). For purposes of the experiment it should be a variable for which the value can be measured; for practical purposes, it should be a variable for which changes in value are of interest to the administrator (or others).

While the method can be used anywhere on the EXPLORATORY ←→ A PRIORI Dimension, more certain (credible) results will be achieved if the administrator's understanding of the circumstances allows him to significantly pre-

determine the effect of the new data on his results (such as by hypothesis testing). This will require identifying those other variables (parameters) which may confound his results by introducing plausible rival explanations (hypotheses), and controlling them by fixing their values, by randomizing their appearance, or by measuring them.

There are several methods for dealing with problems which may be too complex in terms of the desired dependent variable or undesired parameters. One method is to narrow the problem by limiting the number of different changes (independent variables) which will be manipulated at one time. Another method is to "shorten the chain" between the independent and dependent variables by choosing a dependent variable closer in the "causal chain" to the independent variable; this is often possible where the relation between the proximal (near) dependent variable and the desired (distal, or more distant) dependent variable is one that is considered "known" for some practical purpose. A third method is to isolate, as a sub-experiment, that part of a larger experiment which can be defined more clearly and for which the control necessary for hypothesis testing can be realized. Several examples of this simplifying process are provided in Planck (1970).

Where the administrator does not "choose" the problem, he may still have sufficient flexibility in carrying out the experiment to follow some or all of these considerations.

Experimental Design

There are a variety of experimental designs which have been developed in the laboratory and the field. The classic designs which include random assignment of subjects to the experimental and control groups are often difficult to realize in a real world setting. Other designs, less rigorous

but more appropriate for use in some settings, have been developed, and these can be used with a proper awareness of the increased risk to validity (Campbell and Stanley, 1963).

As is true in field testing, the design should include as thorough and complete specification of the methods of collecting data as possible (EXPLORATORY ⟷ A PRIORI Dimension). These methods may include observation, records, interviews and questionnnaires, varying in usefulness according to the variable (or parameter) of interest, and varying in the potential for unintended interaction (STUDY ⟷ EXPERIMENT Dimension).

Similarly, the design should include as thorough and complete specification of the process of data analysis and evaluation as possible. In addition to the reasons previously discussed, the early detailing of this step provides an opportunity to check all of the previous steps. Particularly, the administrator (experimenter) can determine whether or not the expected result(s) will be achieved (and be credible) if he succeeds in carrying out his experiment as planned. By using simulated data, e.g., using values and distributions of the variables that he thinks he can reasonably expect to occur, he can determine if the manipulation he proposes will produce a sufficient range and distribution of values in his variables, whether or not he has provided adequately for the measuring of both variables and parameters, and whether suitable analytical methods (statistics) are available for the kind and amount of data he expects to obtain.

Related Design Considerations

Where possible, the administrator should consider a pilot test of his design to check out its feasibility. The methods of measurement used, especially those which require the cooperation of others, may, if not initially well-designed, create misunderstandings which will result in

inaccurate data on the variables of interest. The pilot data itself may provide a check on the assumptions which the researcher made about the range and distribution of the expected data.

Where information will be obtained which may affect the interests of those furnishing the data or of others in the organization, suitable provisions must be made to ensure privacy and confidentiality. Coding of instruments, assignment of code numbers to disguise the identity of interviewees, and other techniques are available.

A useful practice is to date all drafts and notes so that the record of the experiment can be reconstructed chronologically. Where a document does not in itself contain enough information for later identification, a brief description should be included.

SOME EXAMPLES OF STUDIES AND EXPERIMENTS

For purposes of comparison several different experiences in which the authors participated will be presented. The first three were largely or solely designed and directed by the researcher; the next two involved a considerable degree of participation by administrators; the last one was designed and directed largely by administrators — an "administrative experiment."

Information System Laboratory Experiment

This information system laboratory experiment was carried out at the IBM Research Laboratories. The purpose of the experiment was to evaluate an invention by Peter Luhn, the auto abstract, with a series of competitive concepts including pseudo-auto abstracts, the use of titles, and texts. The procedure involved a typical population, a group of college sophomores from Kings College. They were pre-tested and matched in terms of reading ability, and then given a pre-test which called for knowledge of the subject matter of the material to be used in the experiment. This test became the objective of the information searches, and students were told that they were to answer these questions and have them available during different steps of the experiment. The first step was to divide them up into groups which were then given abstracts, pseudo-abstracts, titles or the full text of the materials which included those upon which the questions were based. They were asked to evaluate the materials in terms of their applicability and to attempt to answer the questions. In the second step, they were allowed to use whatever surrogate they had and, in addition, were allowed to use the text of the full article, and they were again given a chance to answer questions and determine the relevance. The result of

this experiment showed that reasonably good results could be achieved by the pseudo-auto abstracts and the auto abstracts, and that the texts in fact contained much more information even though the subjects did not really have time to read them all. Skimming the text seemed to be more useful than the abstract; the use of the abstract should be relegated to those places where the text is not available.

The experiment was interesting and reasonably compelling except that the translation of the situation of answering a series of questions and sorting documents into a stack of relevant and irrelevant is a far cry from the real use of these surrogates in a real life; and, therefore, the practical import of such a system would involve a field evaluation or initial experiment of a system with auto abstract to allow enough information for making a decision on a policy basis (Rath, Resnick and Savage, 1961).

Document User "Natural Experiment"

This "natural experiment" was designed to determine whether the presence of an abstract in a document which comes to the desk of a scientist or engineer would result in faster and better initial screening decisions. Eighty-five persons in three military laboratories were selected, and records of the decisions made on nearly one thousand documents over a four week period were recorded. By taking advantage of the "natural queue" of documents, it was not necessary to intervene, and this also reduced the amount of unwanted interaction. Other data was obtained by questionnaires, interviews, observation, and use of records. The presence of an abstract did not appear to be a significant factor in the initial screening process (Thompson, 1969a).

Medical Information Field Experiment

A field experiment was designed and carried out in six hospitals in the Chicago area. The purpose of this experiment was to determine the reaction of medical researchers in oncology and cardiology to the introduction of a free, novel, high performance information system. The system consisted of a facsimile terminal and a telephone. They were told they could ask for any library service they needed including searches of literature, inquiries, xeroxes and anything else. The system would deliver to them documents as requested either directly or by mail. Twelve groups participated in the experiment. Half the groups were given the treatment in the first three month period and the other half in the second three month period. All six hospitals used the systems simultaneously. Records were kept of the use of the system, and pre- and post-interviews were carried out. Such a field experiment is quite expensive in that it involves the hiring of personnel, hiring of facsimile equipment, telephone lines, a xerox machine and a great deal of supporting services, and design of interviews and questionnaires. It covered a long period of time and came up with a set of results which were interesting but had been previously expected. The importance was the magnitude of the verification of these results. Such a controlled experiment did develop a substantial amount of data regarding the information system habits and suggested further areas of inquiry to explain the variations. In many cases people were in favor of innovation and having such a system but never used it, giving the excuse they were too busy. Others used it a little, and one or two used the system a lot.

It would be expected, if one listens to discussions, about the information explosion, that a novel free information system would be desired

by the researchers. Offsetting this was the argument (or theory) that information searching habits are developed much earlier in the life of the researcher and a new system would not generally alter his habits. Generally speaking if they used such a novel system it would be when entering new areas where their old habits probably would not be very useful. It was also expected that individual differences would play a key part in the use of such a system. The experiment was carried out successfully as an activity but leaves one feeling the necessity of carrying out many, many variations to answer many of the small points which were brought up. But to carry out those variations would involve a prohibitively large program for a small payoff (Rath and Werner, 1967).

Rehabilitation Counseling Post Hoc Evaluation

This post hoc evaluation involved the study of data gathered on patients who receive vocation counseling at the Rehabilitation Institute. This included looking at the factors used in predicting how well a person would be placed and carry out his job as well as gathering data on what happened. Through the use of factor analysis the study indicated that in fact the predictions were reasonably good regarding the success or not of the patient in the vocational area. Many questions were still left unanswered at the end of the study, but the hospital was willing to accept this data in order to try another evaluation. The problem with post hoc evaluations is that one would have much more confidence in one in which the criteria, the objectives, are established before the evaluation was carried out (Anderson, 1972).

Physical Therapy Simulation

A simulation was carried out of the Physical Therapy Department of the Rehabilitation Hospital. This involved looking at the units of work, equipment and jobs, the patients taken from about three thousand patient data bank, and a detailed study of facilities and activities in the Department. Such research is part of a larger project in systems analysis of the hospital. It was carried out actively with participation of the management of the hospitals, and the cooperation of therapists and every one else involved. The predictions of the simulation model were then compared and checked against data gathered in the Department to develop confidence in the reliability and the validity of the simulation. Once this had been established, predictions were made for a new building the hospital is moving into, and decisions regarding the establishment and the use of resources were based on this. Management was willing to accept the simulation as an input to decision making and acted using these inputs (McKillop and Kennedy, 1971).

Chicago Police Experiment

This experiment was part of a larger program conducted by the Chicago Police Department to evaluate various ways of improving the effectiveness of the department. The objective was to determine if a change in the method of assigning patrol cars would result in an increase in the number of arrests and a decrease in reported crime. In one district, part of the cars assigned to regular beats were relieved of the duty of responding to calls assigned by the communications center and were directed to carry out a continuous, aggressive patrol. The theory to be tested was that extended, uninterrupted periods of patrol would increase the ability of the police officers to observe and apprehend in certain types of crimes, such as auto thefts and burglaries.

Data on the beat cars — time of operation, number of arrests, etc. — was available from the records of the central communication system. A manual log was used to obtain comparable data on the "aggressive patrol" cars. Data on crimes reported was available in police department records.

During the limited experimental period several related hypotheses were tested, but the result of this experiment was more exploratory than hypothesis testing because of the large number of parameters which were discovered (Thompson and Roth, 1970).

Comments

In the first three examples, the researcher was able to predetermine, to a considerable degree, the effect of the new data upon his results (EXPLORATORY ↔ EX POST FACTO Dimension). This might be expected in a laboratory experiment, and, where the purpose of the researcher is to test specific hypotheses, it can also be done in field experiments if the circumstances are not too complex and the design is carefully done. In the next two examples the purpose was, to some degree, exploratory (or evaluative); and in the last, the complex circumstances limited the results to an exploratory level.

All six examples were directly, or nearly so, based on data from the phenomenon (NORMATIVE ↔ EMPIRICAL Dimension). The simulation and post hoc (after-the-fact) evaluation were the most indirect, the first by choice, the second by necessity. The first three, reflecting the purpose of the researcher, were limited to relatively simple experiments, and this simplified the evaluation.

With respect to the STUDY ↔ EXPERIMENT Dimension, the sharpest distinctions are found. The simulation, post hoc evaluation, and "natural

experiment" all fall on the STUDY side of this dimension. All three minimize the possibility of an intended or unintended interaction with the phenomena. In the case of the "natural experiment," the desired "manipulation" occurred without intervention by the researcher. On the EXPERIMENT side, the laboratory experiment would be the most extreme, followed by the field experiment and then the administrative experiment.

This rough categorization necessarily generalizes, or characterizes, or "averages" the various parts of the study or experiment. It would seem clear that a study or experiment could be factored into sub-activities which might then be dimensioned separately. For example, let us take the Document Uses "Natural Experiment." The major thrust, and purpose, was the testing of two detailed a priori hypotheses. In order to operate at the most extreme right of the EXPLORATORY ↔ A PRIORI Dimension, the complete design, including the conclusions for the alternatives, was specified before the first new data was collected. In contrast, the control of rival hypotheses required the gathering of considerable data on processes, but these data were not the central concern of the research. While these may be of separate interest, properly, they should be treated as "additional findings, characteristic of exploratory research," and reported separately (Thompson, 1967, p. 26).

Similarly, on the NORMATIVE ↔ EMPIRICAL Dimension, we can find sub-activities in the course of this field experiment, such as model building and literature search, which are necessary parts of the overall activity. Few, if any, activities, other than the most simple, would not include some sub-activities on the NORMATIVE side, although there may be activities which include little or no EMPIRICAL components. If the sequence of sub-activities is evaluated, it may be suggested that moving from the left to

the right is characteristic of A PRIORI, and from right to left, of EXPLORATORY.

On the STUDY <--> EXPERIMENT Dimension, the central manipulation was a "natural experiment," intentionally avoiding any disruption of the phenomena. Supporting this were use of records, questionnaires, some observations, and interviews ranging from closed to projective. Had there been a manipulation, beyond that intervention necessary to obtain the data, activities over most of the dimension would have occurred.

In principle, it may be possible (and desirable) to develop a more rigorous and detailed definition of the several dimensions. For this purpose a decision-centered framework (spatio-temporal information model) and an input-output (or before-and-after) model of the researcher's purpose might serve. With such a detailed model, specific experiments, evaluations or field tests could be analyzed to determine whether there are patterns of activities which are more "successful" than others, whatever the names used to describe them. If this were true, it might provide an additional basis for the planning of activities.

Without this more detailed definition, the brief discussion of these six examples illustrates the similarities and differences which can be developed through the use of the three dimensions. It would seem clear that a similar analysis can be extended to field testing and evaluation.

SUMMARY

Administrative experimentation can best be described as a specialized form of field testing, field experimentation or evaluation which emphasizes the dual role of administrator and experimenter. The relation among these, the advantages and disadvantages of the method, an outline of the method, and some examples, are discussed.

REFERENCES

Anderson, J. K. A System Analysis Evaluation Framework. Unpublished master's thesis, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, June 1972.

Campbell, D. T. Administrative Experimentation, Institutional Records, and Nonreactive Measures. In J. C. Stanley (Ed.), Improving Experimental Design and Statistical Analysis. Chicago: Rand McNally, 1967, pp. 257-291.

Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963

Caro, F. G. (Ed.). Readings in Evaluation Research. New York: Russell Sage Foundation, 1971.

Chapanis, A. Research Techniques in Human Engineering. Baltimore: The Johns Hopkins Press, 1959.

Ericksen, R.L. and Neal, A.S. Laboratory v.s. Field Experimentation in Human Factors — An Evaluation of an Experimental Self-Service Airline Ticket Vendor. Human Factors, Vol. 13, No. 6, December 1971, pp. 521-531.

Knowles, W. B., Burger, W. J., Mitchell, M. B., Manifan, D. T. and Wulfeck, J. W. Models, Measures, and Judgments in System Design. Human Factors, Vol. 11, No. 6, December 1969, pp. 577-590.

Liebow, E. Tally's Corner. Boston: Little, Brown and Company, 1967.

McKillop, A. E. and Kennedy, O. C. Application of Systems Analysis Programs to the Delivery of Rehabilitation Health Care Services, Final Report. Chicago: The Rehabilitation Institute of Chicago, June 1971.

Morgan, D.L. Evaluation: A Semantic Dilemma, Educational Technology. Vol. 11, No. 12, December 1971, pp. 46-48.

Planek, T.W. Planning Techniques for the Evaluation of Community Support Efficiency in Safety. Human Factors, Vol. 12, No. 1, 1970, pp47-54.

Rath, G. J., Resnick, A. and Savage, T. R. Comparisons of Four Types of Lexical Indicators of Content. American Documentation, Vol. 12, No. 2, 1961, pp. 126-130.

Rath, G. J. and Werner, D. J. "Infosearch: Studying the Remote Use of Libraries by Medical Researchers," The Proceedings of ADI 1967 (American Documentation Institute National Meeting, October 22-27, 1967, New York City), Vol. 4, pp. 58-62, Thompson Book Company, Washington, D. C., 1967.

Suchman, E. A. Evaluative Research. New York: Russell Sage Foundation, 1967.

Thompson, C. W. N. The Initial Screening of Technical Documents by the User, Northwestern University, Department of Industrial Engineering and Management Sciences, June 1969. (a)

Thompson, C. W. N. Management Research Methods for Engineering Managers, SAE Monograph No. 690362, Society of Automotive Engineers, Inc., New York, 1969, 8 pp. (b)

Thompson, C. W. N. Study of the User's 'Natural Queue' of Documents. IEEE Transactions on Engineering Writing and Speech, EWS-13, 1970, pp. 73-78.

Thompson, C. W. N. and Rath, G. J. The Design and Evaluation of Police Experiments. In Proceedings of the 3rd National Symposium on Law Enforcement Science and Technology. Chicago: Law Enforcement Science and Technology Center, IIT Research Institute, 1970.

BEST COPY AVAILABLE

THE DEGREE TO WHICH THE RESEARCHER PREDETERMINES
THE EFFECT OF NEW DATA ON HIS RESULTS

Not at all — Completely

Exploratory Research — Serendipity — Action Research — Evaluative Research — Descriptive and Explanatory Research

Hypothesis Testing — Replication

Systems Design — System Demonstration or Evaluation (Pilot or Field Testing)

Breadboard Testing — Development Testing — Acceptance Testing

Trial and Error — Rote learning — Conditioned Reflex — Imprinting

Flash of insight — Heuristic Models — Preliminary Hypotheses (Mental Models) — Mathematical Models — Algorithmic Models

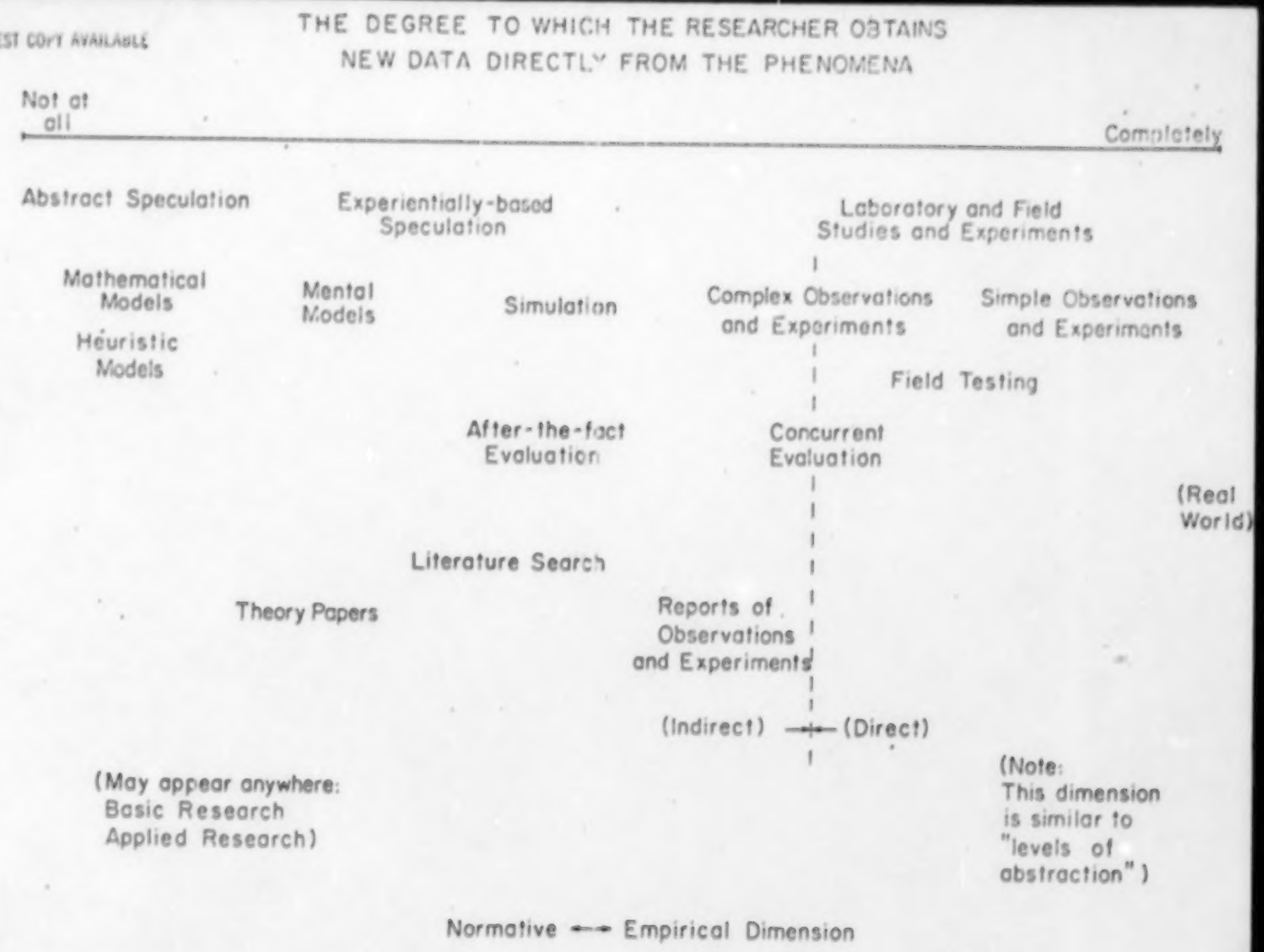
(May appear anywhere:
Evaluation
Field Testing
Field Studies and Experiments
Basic Research
Applied Research
Simulation)

(Data from external stimuli)

(Data from reflection)

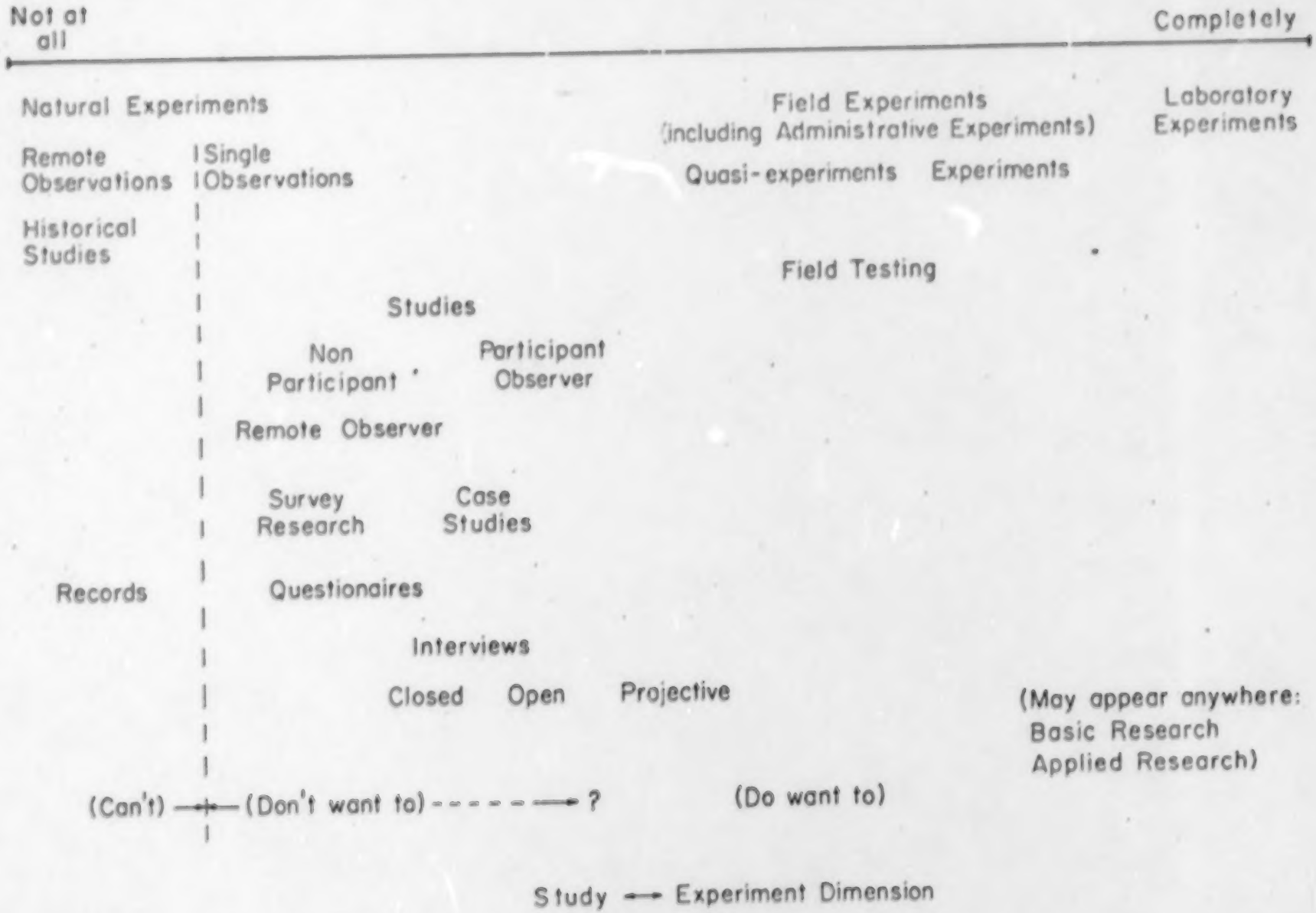
Exploratory ⟷ A Priori Dimension

BEST COPY AVAILABLE



BEST COPY AVAILABLE

THE DEGREE TO WHICH THE RESEARCHER MANIPULATES THE PHENOMENA UNDER OBSERVATION



END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

AUG 27 74